

Dangers of Percentages Proficient: Analysis of Interpretations of High-Stakes Assessment Results Reported on the New Jersey School Report Card

Thomas Tramaglini

Freehold Borough School District, NJ

Structured Abstract

Background: The No Child Left Behind Act of 2001 (NCLB) mandated that all state governments report annual state assessment results in the School Report Card. In New Jersey, the report card provides a snapshot of different input and output variables including comparative assessment data from the state's high-stakes tests. Community members, education leaders, policy makers and boards of education use the School Report Card to judge the quality of their schools and compare them to schools in neighboring districts. The judgments community members make from the reported data can have significant influence on policy decisions that affect public school referenda, elections, and school budgets.

Purpose: Evaluate the use of "percent proficient" (NJDOE, 2006) as the statistical measure for accountability used to report Adequate Yearly Progress (AYP) mandated by the No Child Left Behind Act of 2001. The argument posed is that under the current volatile accountability system (i.e., NCLB), percentage proficient is a weak and inadequate reporting tool. Further, two alternatives for reporting school academic achievement from high-stakes assessments may offer a more accurate snapshot of how public schools are meeting accountability standards.

Research Design: Analytic essay.

Data Collection and Analysis: New Jersey School Report Card data from three school districts in central New Jersey. Analyzed the relationship between "percent proficient" reported on the Report Card and the underlying scale-scores for each district.

Finding and Results: Practical and statistical significance yield two different explanations when one interprets percentages of proficiency. In 2004, 84.5% of Central Middle School (pseudonym) students scored proficient or above on the mathematics Grade Eight Proficiency Assessment (GEPA). In 2005, 79.4% of the students scored proficient or above on the mathematics GEPA. To the layperson, student assessment scores declined 5.1%. This decline may invite complaints by the community and pressure on school district leadership to raise

assessment scores. However, this change may be practically significant and not educationally important. Making a determination of quality using practical significance is less accurate than using a mathematical test of significance to determine if observed effects are noteworthy.

Conclusions: Educators need to use a more accurate statistic of reporting that tracks educational progress that shows more clearly what is really going on in schools. I propose consideration of two alternatives to “percentage proficient” (NJDOE, 2006) for school accountability .

Keywords: school report card, growth model, high-stakes testing

Introduction

High-stakes assessments and their results are an institutionalized component of public schools and their communities. Federal and state governments rely on the assessment results to indicate the quality of education that an individual school or school district delivers. The No Child Left Behind Act of 2001 (NCLB) mandated that all state governments report annual assessment results in the School Report Card. In New Jersey, the report card provides a snapshot of different input and output variables including comparative assessment data from statewide high-stakes tests. Community members, education leaders, policy makers, and boards of education use the New Jersey (NJ) School Report Card to judge the quality of their schools and compare them to schools in neighboring districts. The judgments community members make from the reported data can have significant influence on policy decisions that affect public school referenda, elections, and school budgets. Should the public trust the salience of the data reported on the School Report Card?

Using real data from a school district in central New Jersey as context, I evaluated the use of “percent proficient” (NJDOE, 2006) as the statistical measure for accountability used to report Adequate Yearly Progress (AYP) under NCLB. The argument posed is that under the current volatile accountability system (i.e., NCLB), percent proficient is a weak and inadequate reporting tool. Further, two alternatives for reporting school academic achievement on high-stakes assessments may offer a more accurate snapshot of how public schools are meeting accountability standards.

Problem

In the United States and NJ, high-stakes testing is the hallmark of the current school accountability scheme. Whether the National

Assessment of Education Progress (NAEP), American College Testing Program (ACT), SAT Reasoning Test (SAT), New Jersey Assessment of Skills and Knowledge (NJ ASK), or the High School Proficiency Assessment (HSPA) is the tool for measurement, people in NJ use the testing results from those assessments as performance indicators for schools and/or districts. Schools and communities make decisions about educational quality based on the percentage of students that are or are not proficient on high-stakes tests. However, interpreting academic achievement using percentage proficient may not be an accurate method of determining academic performance. That is, statistical reports made by educational bureaucrats may present a superficial model of reporting high-stakes assessment results. What is the difference between the results reported publicly on the NJ School Report Card and other models that might be better indicators of academic performance?

Case: Central School District (Pseudonym)

“Central Township” is a suburban school district in central New Jersey. Over the past ten years, the school district grew from a one-school K-8 district of 990 students to a four-school, K-12 school district enrolling over 2,000 students. Until 2005-2006, the community trusted the school district to provide a quality education. Residents passed budgets and referenda without resistance. Parents believed in initiatives that moved the district towards academic achievement on standardized assessments.

The town’s socioeconomic status is higher than it was 10 years ago. With the influx of a more educated and wealthy constituency, the focus of the community toward the schools changed. The town’s residents now place more emphasis on results of high-stakes assessments than they did a decade ago. Once a relatively passive township, now the residents of Central examine and interpret assessment scores on the NJ School Report Card. Two years ago, parents and community members began to question the school on what they perceived to be declining assessment scores and the district’s perceived inability to improve academic achievement on the NJ ASK 3 & 4, GEPA, and the HSPA.

Parental complaints about student assessment scores followed the release of the 2004 NJ School Report Card. Tables 1 and 2 present the Central School District school report card information provided by the New Jersey Department of Education (NJDOE). The tables highlight the NJ ASK3 scores reported on the 2005-2006 New Jersey School Report Card. The report card presents the percentages of students rated partially proficient, proficient, and advanced proficient. Presumably,

data allow readers to compare student progress over two years and against proficiencies of the state and districts in similar factor groups.

Table 1

New Jersey Assessment of Skills and Knowledge (NJASK3) – Language Arts Literacy^a

				<u>Proficiency Percentages</u>		
		Year	<i>n</i>	Partial	Proficient	Advanced
All Students	School	2004-2005	154	32.1%	63.4%	4.5%
		2003-2004	131	30.5%	68.0%	1.6%
	District	2004-2005	154	32.1%	63.4%	4.5%
		2003-2004	131	30.5%	68.0%	1.6%
	DFG	2004-2005	12,697	13.7%	82.5%	3.8%
		2003-2004	×	×	×	×
	State	2004-2005	100,931	16.7%	78.8%	4.4%
		2003-2004	103,994	20.7%	75.6%	3.8%

^aTo protect the privacy of students, the Department of Education suppresses sufficient information to eliminate the possibility that personally identifiable information will be disclosed.

× The DFG data for 2003-04 were omitted because they were based on the 1990 Census. For 2004-05, the source for DFG data is the 2000 Census.

Source. New Jersey Department of Education School Report Card, 2005

Similar to the other assessments in the district, School Report Card data for the Grade 3 NJ ASK demonstrated what appeared to be a drop in assessment scores. Community members did not accept the district’s suggestion that the decline in scores could have been caused by the inclusion of special education students into the general population of test scores. No longer were assessment scores higher than 90% proficient and advanced proficient. Parents and residents in Central Township reacted angrily to the perceived decrease in students’ proficiency in mathematics and language arts.

Table 2

New Jersey Assessment of Skills and Knowledge (NJASK3) – Mathematics^a

				<u>Proficiency Percentages</u>		
		Year	<i>n</i>	Partial	Proficient	Advanced
All Students	School	2004-2005	154	36.6%	50.4%	13.0%
		2003-2004	131	29.2%	66.2%	4.6%
	District	2004-2005	154	36.6%	50.4%	13.0%
		2003-2004	131	29.2%	66.2%	4.6%
	DFG	2004-2005	12,742	14.8%	56.5%	28.7%
		2003-2004	×	×	×	×
	State	2004-2005	101,683	17.5%	53.7%	28.8%
		2003-2004	103,812	23.4%	53.8%	22.8%

^aTo protect the privacy of students, the Department of Education suppresses sufficient information to eliminate the possibility that personally identifiable information will be disclosed.

× The DFG data for 2003-04 were omitted because they were based on the 1990 Census. For 2004-05, the source for DFG data is the 2000 Census.

Source. New Jersey Department of Education School Report Card, 2005

Understanding Percentage Proficiency

Under the provisions of the No Child Left Behind Act of 2001, AYP is measured in percentage proficient. In New Jersey, the state scores its assessments using scale scores that range from 100-300. A score of 200 represents the cut-score for determining if a student is proficient or partially proficient. Percentage proficient indicates the percentage of students who scored at or above the cut-score for proficiency (i.e., 200). To the layperson, this seems easy to understand, but several weaknesses exist.

Variability

A weakness of using percentage proficient is the failure of *percentage* to determine accurate growth and decline estimates in mean student achievement. Proficiency percentages miss a great deal of what occurs when mean assessment scores rise or decrease (Choi, Goldschmidt, & Yamashiro, 2005). For instance, in 2004 the mean statewide proficiency percentage of Central District students on the

mathematics section of the NJ ASK3 was 66.2%. In 2005, the mean percentage had fallen to 50.4% of the students scoring at the proficient cut score. However, the two percentages do not reflect the frequencies and/or distributions of students scoring above or below the cut score. That is, percentages of proficiency do not offer detailed information about the variability of the statistic.

The variability of a statistic indicates the spread of its sampling distribution (Moore & McCabe, 2006). To readers, 66.2% proficiency means that nearly 1 in 3 children did not pass the assessment. This may not be true, especially in small populations where a few children could skew significantly the scores. Not having the complete picture of academic achievement might cause education leaders and community members to make faulty decisions or apply misguided remedies (Ramaley, 2005). Without knowing the variability associated with test scores, readers' interpretations may lead to false negative or false positive errors.

Reliability of Assessment Scores

Researchers argued that the results of most high-stakes assessments are not reliable measures of academic achievement (Linn, 2005; Ramaley, 2005; Rogosa, 2005; Tienken & Wilson, 2005). In turn, the interpretations that parents and community members make from high-stakes assessments may fall victim to poor conclusion validity. If assessment results lack reliability, then the conclusions that follow may be flawed. This is a reality in New Jersey. Examining the technical characteristics of large-scale state assessments of skills and knowledge, Tienken and Wilson (2005) indicated that the use of assessment instruments such as the GEPA to determine academic achievement do not represent a complete measure of mandated content standards (content validity) and the test results lack reliability that must accompany a high-stakes accountability system.

Standard Error of Measurement

The measure of the variability of the assessment score may explain why some of these interpretations are unreliable. Standard error of measurement (Gall et al., 2003) allows a determination to be made of the probable range within which true scores fall (Moore & McCabe, 2006). Consider the standard error of measurement for the 2004 GEPA (Table 3). The standard error of measurement for scale scores is 13.02 points on the mathematics section of the GEPA. Therefore, if the mean

score were to fluctuate more than 26 points, the class measure will not be a very accurate calculation of academic achievement. Consequently, interpretations made by laypeople based on the reporting of scores that exhibit large standard errors, such as those from the GEPA, may be flawed and inaccurate. Similar interpretation made by policy makers may send an inaccurate message to the community at large.

Size Effect and Inequity of Assessment Scores

Report card data purportedly provide a comparative analysis to determine how a particular school performs on assessments in comparison to other schools, districts, and AYP requirements. However, a significant weakness in reporting assessment scores using proficiency percentages is the failure of percentages to identify that differences in sample sizes may cloud comparative relationships among the samples. Although the state publishes the number (*n*) of students who completed the assessment on the report cards, I argue whether many parents, for example, possess the statistical knowledge to acknowledge and understand the influence sampling size can have on the average test score. The assumption is that the majority of people who make inferences from school report card data do not consider the influence of sample size when they make decisions about assessment scores.

Table 3

Standard Error of Measurement for the 2004 GEPA

Standard Error of Measurement – 2004 GEPA			
	<u>Language Arts</u>	<u>Mathematics</u>	<u>Science</u>
GEPA	12.43	13.02	10.80

Source: NJDOE, GEPA 2004 Technical Report

One common way that the State of New Jersey presents data on the NJ State Report Card, which the local newspapers publish, is by comparing trend data from year to year. Table 4 is an example of published comparison data from Central School District. The data represent district scores from 2004 and 2005 on the NJ ASK 4. In 2004, 78.6% of the students in this school’s fourth grade scored proficient or above in mathematics; 73.1% scored proficient or advanced proficient in 2005.

Table 4

Central School District NJ ASK4 Percent Proficient for 2004 and 2005

Mathematics						
2004 Percentage Proficient				2005 Percentage Proficient		
School	Partially Proficient	Proficient	Advanced Proficient	Partially Proficient	Proficient	Advanced Proficient
Sample Elementary School	21.4%	76.1%	2.5%	27.0%	68.9%	4.1%

Parents and community members in Central District would be appalled by a 5.5% decline in fourth-grade math scores. However, the hidden statistic in this table is sample size. In 2004, the fourth grade had 154 students. In 2005, the fourth grade had only 113. The difference between 154 students and 113 students in the case for percent proficient is significant: 1.36. That is, students in 2005 counted nearly 1.36 more times toward proficiency than the students did in 2004. The difference limits the accuracy of the statistic.

The layperson usually does not consider sample size when comparing assessment scores between districts. Table 5 compares Central's NJ ASK4 scores to those of similar school districts such as "Masonry" and "The River." The table also includes the mean scores for the district factor group and the state. From the table, Central readers would infer that their fourth grade students did not perform as well as the students in the Masonry and The River school districts. However, differences in school sample sizes could skew the statistical interpretation. If a person were to compare these three school districts, one student's score at Central would affect the percentage proficiency 5.41 more times than would the score of a Masonry student and 9.74 more times than the score of a student at The River. The same statistical inequality would also exist between comparisons of the State and DFG. The inequities in this example are troubling.

Table 5

Comparison of Percent Proficient from Central School District and Similar Districts

<u>Mathematics</u>		<u>Percent Proficient</u>		
School	<i>n</i>	Partially Proficient	Proficient	Advanced Proficient
Central	100	25.7	72.3	2.0
Masonry	541	20.5	73.1	6.4
The River	974	19.2	75.6	5.2
DFG	302*	22.8	72.8	4.4
State	299*	19.2	76.9	3.9

* Represents the \bar{n} in that group*Practical Versus Statistical Significance*

Practical and statistical significance yield two different explanations when one interprets percentages of proficiency. In 2004, 84.5% of Central Middle School students scored proficient or above on the mathematics GEPA. In 2005, 79.4% of the students scored proficient or above on the GEPA. To the layperson, student assessment scores declined 5.1%. This decline may invite complaints by the community and exert pressure on school district leadership to raise assessment scores. However, this change may only be a practically significant one. Practical significance occurs when one looks at changes in assessment scores and the changes appear to be large enough to determine that the change was not due to chance (Gall, Gall, & Borg, 2003). Making a determination using practical significance is less accurate than using a statistical test of significance to determine if observed effects are noteworthy.

Tests of significance determine if observed effects are large enough to preclude the effects occurring by chance (Gall, et al., 2003; Moore & McCabe, 2006). Yet, *Z* and *T* scores never appear in the newspapers or on School Report Cards. *Z* or *T* scores determine the rejection of the null hypothesis. That is, does the 5.1% really mean that academic achievement is declining because of something the district did or did not do, or did the decline happen by chance? Although the percent of Central's students who scored proficient or above on the GEPA is

practically significant, their scores were not statistically significant, meaning that variance in the scores between years was likely to occur by chance. However, people in Central District interpreted the scores as declining and ultimately insisted that the school take action.

Implications for Practice

Community members, parents and other stakeholders can be highly influential in the daily operations of schools and school districts. The significance of high-stakes assessment scores can serve as a catalyst for stakeholders to make important decisions on whether to support the school district's policies and procedures, thus underscoring the importance of an accurate measure of academic achievement in the district. As in the Central case, parents and community members developed conclusions from the assessment scores using the percent proficient statistic. As argued, the implications of using a weak statistic such as percent proficient as the measuring stick for academic achievement can be risky and even detrimental to teaching and learning.

Clearly, educators need to use a more accurate statistic of educational reporting that tracks educational progress and demonstrates what is really going on in schools relative to student achievement. I propose consideration of two alternatives to "percent proficient" (NJDOE, 2006) for school accountability. The first method is the use of scale scores. Scale scores come from student raw scores and allow actual scores to determine how an average group (i.e., class, school, or district) compares to others. The second alternative presents Value-Added Modeling, a system of building composite data comparisons using growth and time to analyze test score increases and declines.

Scale Scores

Unlike percentages, scale scores present a more accurate picture of student achievement. Scale scores represent a conversion of raw scores placed on a scale that allows for averaging across students (Choi, et al., 2005; Linn, 2005). Scale scores derive directly from the students' raw scores and are not a status indicator such as percent proficient. Table 6 presents a comparison of Central's GEPA Science scores in 2004 and 2005. The percentage proficiencies in 2004 and 2005 were both 73.4%. To the layperson, the school did not improve. In this case, scale scores demonstrate that the case is much different. From 2004 to 2005, the average scale score rose 8.6 points. Further analysis conducted using disaggregated sub-scale score or cluster data can explore the *whys* and

hows of this increase. Investigation could explore whether the variability of the scores from one year to another was due to a lower *n*. Currently, New Jersey only reports scale scores to school district assessment coordinators and does not report scale scores to the public.

Performance over Time

As demonstrated, scale scores can explain changes over time more accurately than percentages of proficiency. Another statistical analysis tool that can measure performance over time is the Value-Added Model (VAM). VAM represent a specific subset of accountability models that provide estimates of students’ progress over time (Choi, et al., 2005; Koretz, 2005; Linn, 2005). Using any number of composite designs, VAM chart growth over time and provide an organized comparison of schools. Figure 1, from Choi, et al. (2005) is one example of a VAM that compares schools’ growth over time.

Table 6

Example of Central School District New Jersey GEPA Science Percent Proficient and Scale Scores

Category	<u>2004</u>	<u>2005</u>	<u>04-05 Diff.</u>
Percent Proficient/ Advance Proficient	73.4%	73.4%	0
Scale Score	211.5	220.1	+8.6 points

The benefits of using Value-Added Models are more powerful than using percent proficient because scores identify actual gains and decreases measured over time, whereas in percentages actual scores are not used. Further, the adjustments (such as controlling for SES or weighting) could improve the accuracy of comparing schools to each other. McCaffery, Lockwood, Koretz, Louis, and Hamilton (2004) stated:

Enthusiasm for this approach stems in large part from the belief that it can remove the effects of factors not under the control of the school, such as prior performance and socioeconomic status, and thereby provides a more accurate

indicator of school or teacher effectiveness than is possible when these factors are not controlled. (p. 68)

Using VAM yields advantages over the current practice of reporting percent proficient. VAM can incorporate simple and complex multivariate data in a longitudinal format. That is, researchers can collect data over multiple years and analyze it statistically to estimate the effects on student achievement of individual schools or teachers. For instance, McCaffery, et al. (2004) investigated the use of VAM and concluded that the use of multivariate longitudinal models presents salient descriptions of academic achievement. Other researchers (Bryk, Thum, Easton, & Luppescu, 1998; Choi, Yamashiro, Seltzer, & Hermann, 2004) affirmed that the use of multivariate approaches to analyze student achievement data presented a more accurate description of academic performance.

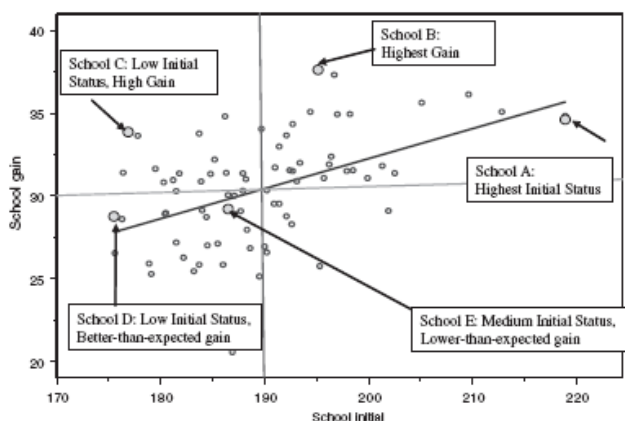


FIGURE 1

Effect of different criteria for determining school performance: school rank by different criteria.

1. Status (x-axis): $A > B > E > C > D$
2. Gain (y-axis): $B > C > A > E > D$
3. Conditional gain (regression line): $B > C > D > E > A$

Note that the vertical line and the horizontal line represent district average initial status and district average gain, respectively.

Figure 1. Sample Value Added Model comparing composite coefficients of school growth. From "Exploring Models of School Performance: From Theory to Practice," by Kilchan Choi, Pete Goldschmidt, and Kyo Yamashiro, 2005, *Uses and Misuses of Data for Educational Accountability and Improvement: The 104th Yearbook of the National Society for the Study of Education: Part 2*, Malden, MA: Blackwell Publishing, pp.119-146.

While growth models (VAM) are a promising avenue for data analysis and reporting, implementing such an approach may have limitations. VAM are less common around the world and most state

departments of education lack the experience necessary to implement them at this time. Growth models rely on consistent longitudinal data through the use of yearly testing in each grade. Yearly testing translates into more time spent testing, higher costs, and reduced instructional time. Finally, another limitation of growth or VAM is the statistical complexity of arriving at conclusions and the challenge of educating the public to understand the data. Board of education members, parents, policy makers, and community members seem to want simple explanations of complicated phenomena. While VAM offer more accurate indicators of school progress than some other indicators, it is likely that the layperson may not understand what is presented. Studies need to overcome these challenges and guide practitioners and policymakers to achieving synergy.

Conclusion

School accountability in the 21st century will likely continue to rely on the industrial production principle (e.g. output variables) that schools must meet increasing academic achievement requirements (most notably AYP). Rogosa (2005) posited, “Designing a useful accountability system requires an understanding of which combinations of measuring and judging best serve the intended policy objectives, while also having defensible statistical properties” (p. 170). The simple fact that the New Jersey Department of Education continues to report assessment scores using percent proficient is perilous, given the inherent flaws in the process and the turbulent educational environment. The inferences that community members would draw, given the reliability measures of the assessment results, may be different when interpreting and placing emphasis on proficiency percentages.

Although No Child Left Behind mandates that schools report percentage proficiencies to the public, a situation exists in which educational policy surpasses the statistical knowledge of the average school stakeholder. Using accurate models of reporting to inform the public of school progress would likely allow better inferences by the community about their schools and a more stable learning community. The use of either scale scores as the statistic of analysis or by integrating some form of Value-Added Model would better serve this purpose. Ultimately, communities play an important role in schools and at a minimum, school communities need accurate representation of the data so their interpretations are salient.

References

- Bryk, A.S., Thum, Y.M., Easton, J.Q., & Luppescu, S. (1998). Accessing school academic productivity: The case of Chicago school reform. *Social and Quasi-Experimental Designs for Research*, 2, 103-142.
- Choi, K., Goldschmidt, P., & Yamashiro, K. (2005). *Exploring models of school performance: From theory to practice*. In J.L. Herman & E.H. Haertel (Eds.), *Uses and Misuses of Data for Educational Accountability and Improvement: The 104th Yearbook of the National Society for the Study of Education* (pp. 119-146). Malden, MA: Blackwell Publishing.
- Choi, K., Yamashiro, K., Seltzer, M. & Herman, J. (2004, September). *Comparing like with like: The role of students and school characteristics in value-added models*. Paper presented at the Center for Research on Evaluation, Standards, and Student Testing (CRESST) Annual Conference, Los Angeles, CA.
- Gall, M.D., Gall, J.P., & Borg, W.R. (2003). *Educational research: An introduction* (7th ed.). Boston, MA: Allyn and Bacon.
- Koretz, D. (2005). Alignment, high stakes, and the inflation of tests scores. In J.L. Herman & E.H. Haertel (Eds.), *Uses and Misuses of Data for Educational Accountability and Improvement: The 104th Yearbook of the National Society for the Study of Education*. (pp. 99-117). Malden, MA: Blackwell Publishing.
- Linn, R.L. (2005). Issues in the Design of Accountability Systems. In J.L. Herman & E.H. Haertel (Eds.), *Uses and Misuses of Data for Educational Accountability and Improvement: The 104th Yearbook of the National Society for the Study of Education*. (pp. 78-98). Malden, MA: Blackwell Publishing.
- McCaffery, D. Lockwood, J.R., Koretz, D., Louis, T., and Hamilton, L. (2004). Models for value-added modeling accountability systems. *Journal of Educational and Behavioral Statistics*, 29(1), 67-101.
- Moore, D.S., & McCabe, G.P. (2006) *Introduction to the practice of statistics*. (5th ed.). New York: W.H. Freeman and Company.
- New Jersey Department of Education (2005). 2004 Grade eight proficiency assessment technical report. #1505.01. NJDOE.
- New Jersey Department of Education. (2006). 2005-2006 School Report Card. Retrieved March 19, 2006: <http://education.state.nj.us/rc/rc05/menu/29-4190.html>.

- Ramaley, J.A. (2005). Goals for learning and assessment. In J.L. Herman & E.H. Haertel (Eds.), *Uses and Misuses of Data for Educational Accountability and Improvement: The 104th Yearbook of the National Society for the Study of Education* (pp. 55-75). Malden, MA: Blackwell Publishing.
- Rogosa, D. (2005). Statistical misunderstandings of the properties of school scores and school accountability. In J.L. Herman & E.H. Haertel (Eds.), *Uses and Misuses of Data for Educational Accountability and Improvement: The 104th Yearbook of the National Society for the Study of Education* (pp. 147-174). Malden, MA: Blackwell Publishing.
- Tienken, C.H., & Wilson, M.J. (2005, January). *Technical characteristics of large-scale state assessments of skills and knowledge*. Paper presented at the annual conference for the New Jersey Association for Supervision and Curriculum Development, East Windsor, NJ.
- U.S. Department of Education. (2002). No Child Left Behind: A Desktop Reference. Retrieved November 22, 2005: <http://www.ed.gov/admins/lead/account/Nclbreference/reference.doc>.

Thomas Tramaglini is the Director of Technology and Assessment for the Freehold Borough School District in Monmouth County, NJ. Correspondence regarding this article should be sent to ttram@freeholdboro.k12.nj.us